

Getting Started with Cloudera Flow Management

Date published: 2019-06-26

Date modified: 2023-08-29

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

What is Cloudera Flow Management?	4
Setting up Cloudera Flow Management	4
Download NiFi and NiFi Registry	5
Install NiFi and NiFi Registry	5
Start NiFi and NiFi Registry	5
Building your first data flow	5
Connecting NiFi to NiFi Registry	9
Creating a bucket in NiFi Registry	9
Versioning data flows	10
Start version control	10
Commit local changes	12
Change the version of your data flow	13
Manage versioned data flows	14

What is Cloudera Flow Management?

Cloudera Flow Management (CFM), powered by Apache NiFi and NiFi Registry, is a key component of the Cloudera DataFlow (CDF) platform. This documentation provides an overview of CFM, its capabilities, and the integration with various components.

Key features and capabilities

Code-free data ingestion

CFM allows you to ingest and manage data without the need for coding, providing a user-friendly solution for data movement within the enterprise.

Versatile data capture

You can capture diverse data types from any source securely, facilitating the seamless flow of data into the enterprise.

Dynamic routing and data transformation

You can transform and dynamically route data in real-time, ensuring flexibility in managing and optimizing data flows.

Complex and scalable data flows

CFM allows you to build intricate and scalable data flows, tailoring them to meet specific business requirements.

Real-time monitoring and management

You can manage, edit, and monitor data flows in real-time, offering transparency and control over the entire data movement process.

DevOps-style flow development lifecycle

You can adopt a DevOps-style development lifecycle, leveraging templates, version control, collaboration, and diverse deployment options for faster application delivery.

Seamless integration

CFM seamlessly integrates with CDF components such as Kafka, MiNiFi, Flink, Spark Streaming, as well as platform components like HBase, Hive, Impala, and Kudu.

Main components

Apache NiFi

- Core data ingestion engine featuring a no-code graphical user interface
- Supports 300+ processors for connectivity, transformation, and content routing

Apache NiFi Registry

- Companion to NiFi enabling DevOps-style flow development
- Supports versioning flows, promoting flows, and deploying flow applications across environments

Setting up Cloudera Flow Management

To create and manage your data flows using Cloudera Flow Management (CFM), you need to download and install two key components: NiFi and NiFi Registry.

Follow these step-by-step instructions to start your journey with CFM.

Download NiFi and NiFi Registry

You can download NiFi and NiFi Registry from the CFM repository. Review the *Download Locations* to identify the appropriate download link suitable for your operating system and operational objectives.

Related Information

[Download locations](#)

Install NiFi and NiFi Registry

For a comprehensive understanding of the Cloudera Flow Management deployment process, review the *CFM installation workflow* documentation. This resource provides insights into the various deployment scenarios available and offers step-by-step installation instructions tailored to your specific requirements.

Related Information

[CFM installation workflow](#)

Start NiFi and NiFi Registry

To start using NiFi and NiFi Registry, open the applications and access their web User Interfaces (UIs). The UIs are designed to be intuitive, simplifying the process of creating data flows with Cloudera Flow Management (CFM).

The NiFi UI offers a wide variety of options for creating, visualizing, editing, monitoring, and administering data flows. To explore its functionalities, see the detailed guidance provided in the *Apache NiFi User Guide*.

The NiFi Registry UI displays the available shared resources and provides tools for creating and administering users/groups, buckets, and policies. Learn more about its features and functionalities in the *Apache NiFi Registry User Guide*.

Related Information

[Apache NiFi User Guide](#)

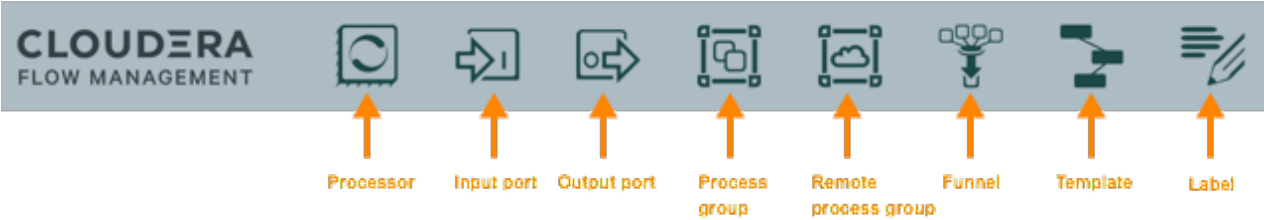
[Apache NiFi Registry User Guide](#)

Building your first data flow

You can use Apache NiFi to create data flows for any protocols or data sources. Start by constructing your flow with a listener component, retrieving data from a source system, and conclude it with a transmitter component that writes the data to a target system. NiFi provides a range of components for processing data efficiently that you can add in between your listener and transmitter.

NiFi component toolbar

The NiFi Component toolbar, located in the top-left corner, contains diverse components that you can use when constructing data flows. These components act as the building blocks, enabling you to design data flows tailored to your specific needs.



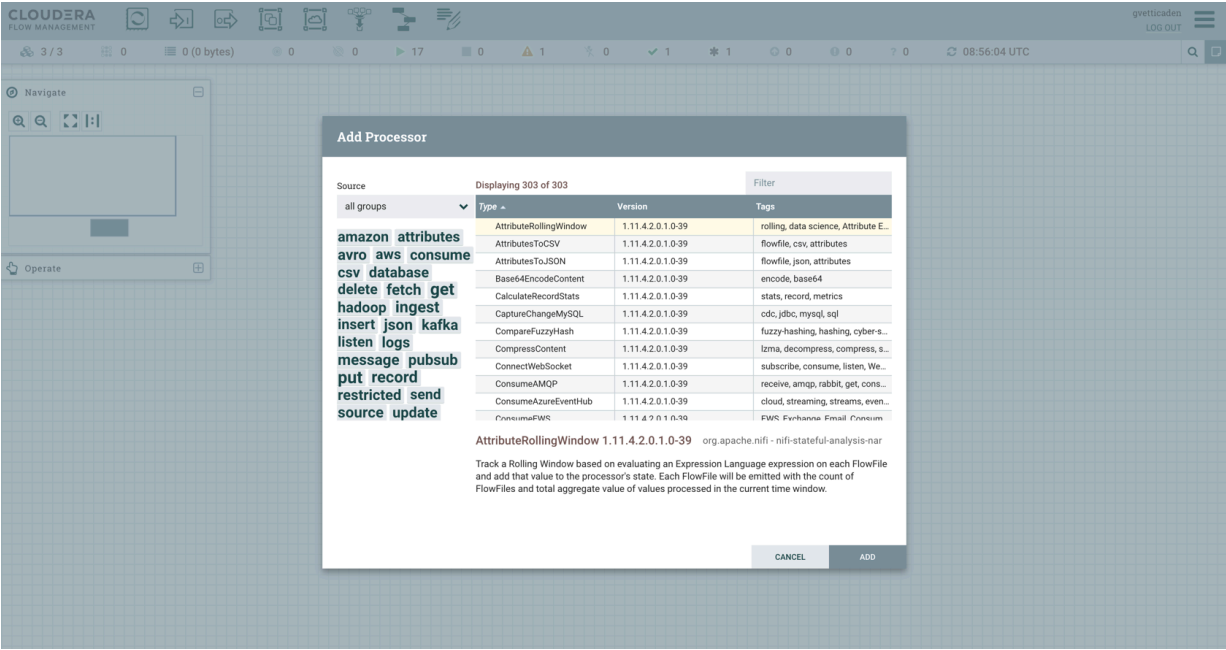
Creating a data flow

About this task

This section walks you through the process of creating NiFi data flows from scratch using basic flow options. For information about building advanced data flows (using parameters, custom properties, controller services, or reporting tasks), see the *Apache NiFi User Guide*.

Procedure


- 1. Add data flow components.
To add a processor to your flow, drag processor icons from the component toolbar to the canvas, located in the center of the screen. This step initiates the creation of your data flow.
For more information, see *Adding Components to the Canvas* in the *Apache NiFi User Guide*.



2. Configure the processors.

- Right-click or double-click a processor on the canvas and select Configure in the context menu.
A configuration dialog with the following tabs is displayed: Settings, Scheduling, Properties, and Comments.
- Configure the processor according to the behavior you desire in your data flow.
- When you have finished configuring the options you need, click Apply to save the changes or cancel all changes by clicking Cancel.

Configure Processor

 Invalid

SETTINGS

SCHEDULING


PROPERTIES

COMMENTS

Name

RouteOnAttribute

☒ Enabled

Automatically Terminate Relationships 
☐ unmatched
FlowFiles that do not match any user-defined expression will be routed here

Id


dd219e19-81a7-1394-0000-00007022e5c2

Type


RouteOnAttribute 1.11.4.2.0.1.0-39

Bundle


org.apache.nifi - nifi-standard-nar


Penalty Duration 

30 sec

Yield Duration 

1 sec

Bulletin Level 

WARN 

CANCEL

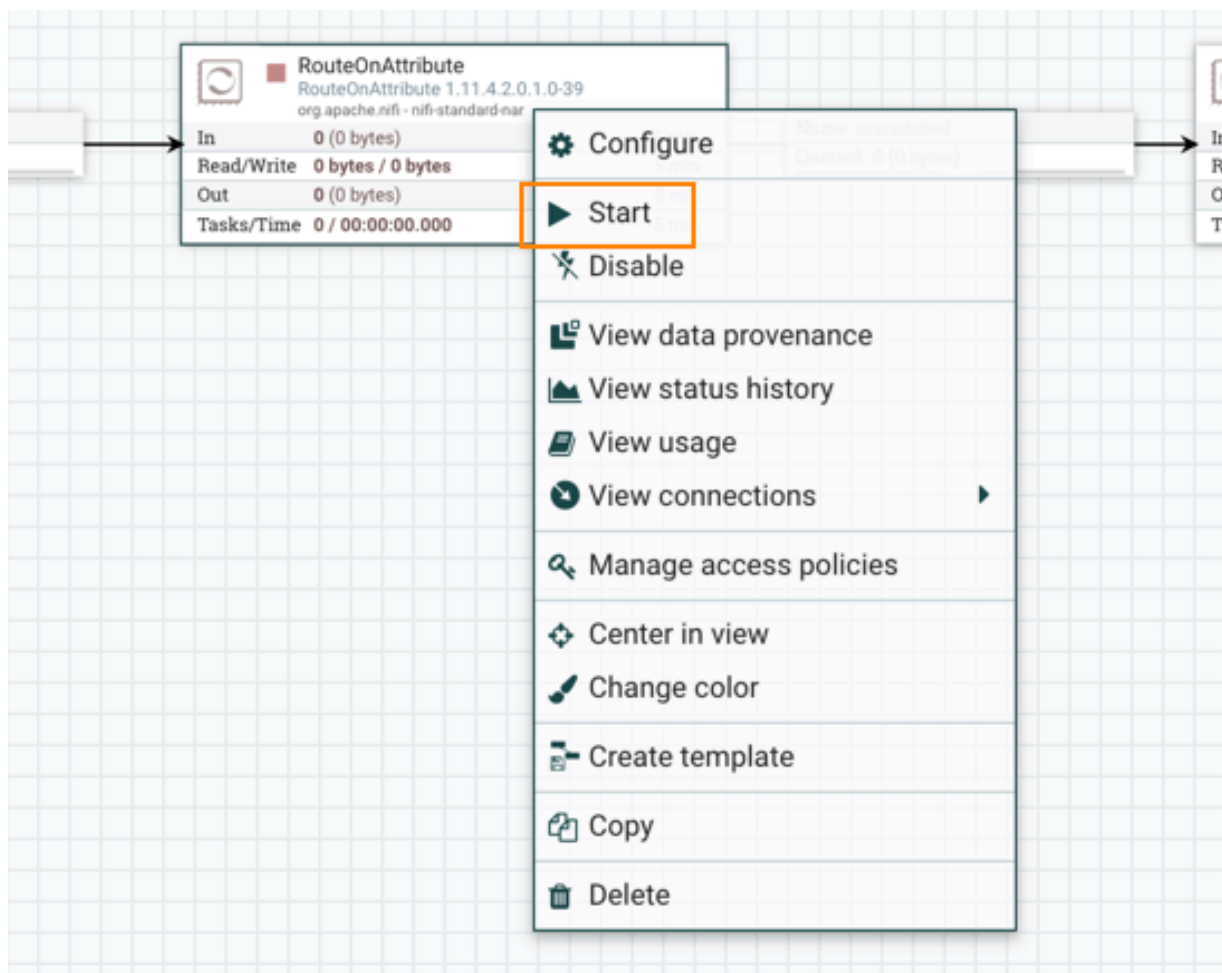
APPLY

3. Connect data flow components.

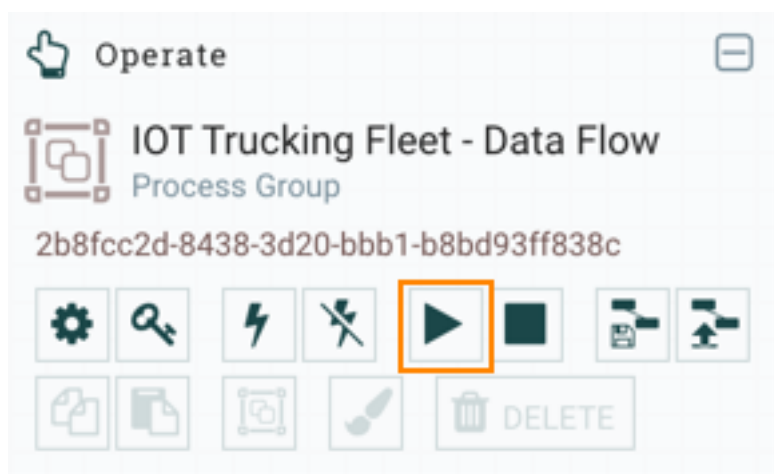
Hover over a component, drag the connection icon to another component, and select the relationships you want to include for this connection.

4. Start your data flow.

Even when components are connected, the flow is inactive. To activate processors, individually click each one, then right-click and choose Start from the menu.



Alternatively, select all processors and click the Start icon on the Operate palette.



After starting the data flow, the processor icons in the top-left corner change from a stopped to a running state. You can stop the processors by using the Stop icon on the Operate palette or the Stop context menu item.

Results

By following these steps, leverage NiFi components to construct a dynamic data flow suitable for your specific use case.

Related Information

Apache NiFi User Guide

Connecting NiFi to NiFi Registry

To enable support for versioned flows in Cloudera Flow Management, follow these steps to configure NiFi and establish a connection with NiFi Registry.

Before you begin

- You have added and configured NiFi and NiFi Registry.
- You have started both NiFi and NiFi Registry.

Procedure

1. On the NiFi UI, select Controller Settings from the Global Menu.
2. Select the Registry Clients tab.
3. Click the Add icon (+) to launch the Add Registry Client dialog.
4. Add the name and URL location of the NiFi Registry service.

Results

By completing these steps, you have established the connection between NiFi and NiFi Registry, enabling support for versioned flows within Cloudera Flow Management. This integration facilitates the effective management and version control of your data flows.

NiFi Settings

GENERAL	REPORTING TASK CONTROLLER SERVICES	REPORTING TASKS	REGISTRY CLIENTS
---------	------------------------------------	-----------------	------------------

Name ▲	Location	Description
NiFi Registry	https://cfm-node-3.cloudera.com:18433	

Creating a bucket in NiFi Registry

NiFi Registry enables you to store and organize versioned data flows efficiently using buckets. You can create multiple buckets to align with your categorization preferences, supporting organization based on teams, deployment environments, usage patterns, or business units.

Procedure

1. Open the NiFi Registry UI.
2. Click the wrench icon (Settings) in the top right corner of the screen.
3. Click New Bucket in the Buckets window.
4. Enter a name for your bucket, reflecting its purpose or categorization.
5. Click Create.

Results

You can see that a new bucket is visible in the NiFi Registry.

For more information on how to manage Registry buckets, see the *Apache NiFi Registry User Guide*.

Related Information

[Apache NiFi Registry User Guide](#)

Versioning data flows

When NiFi is connected to a NiFi Registry, data flows can be version controlled at the process group level. Learn how you can work with different flow versions in NiFi and save them in NiFi Registry.

Before you begin

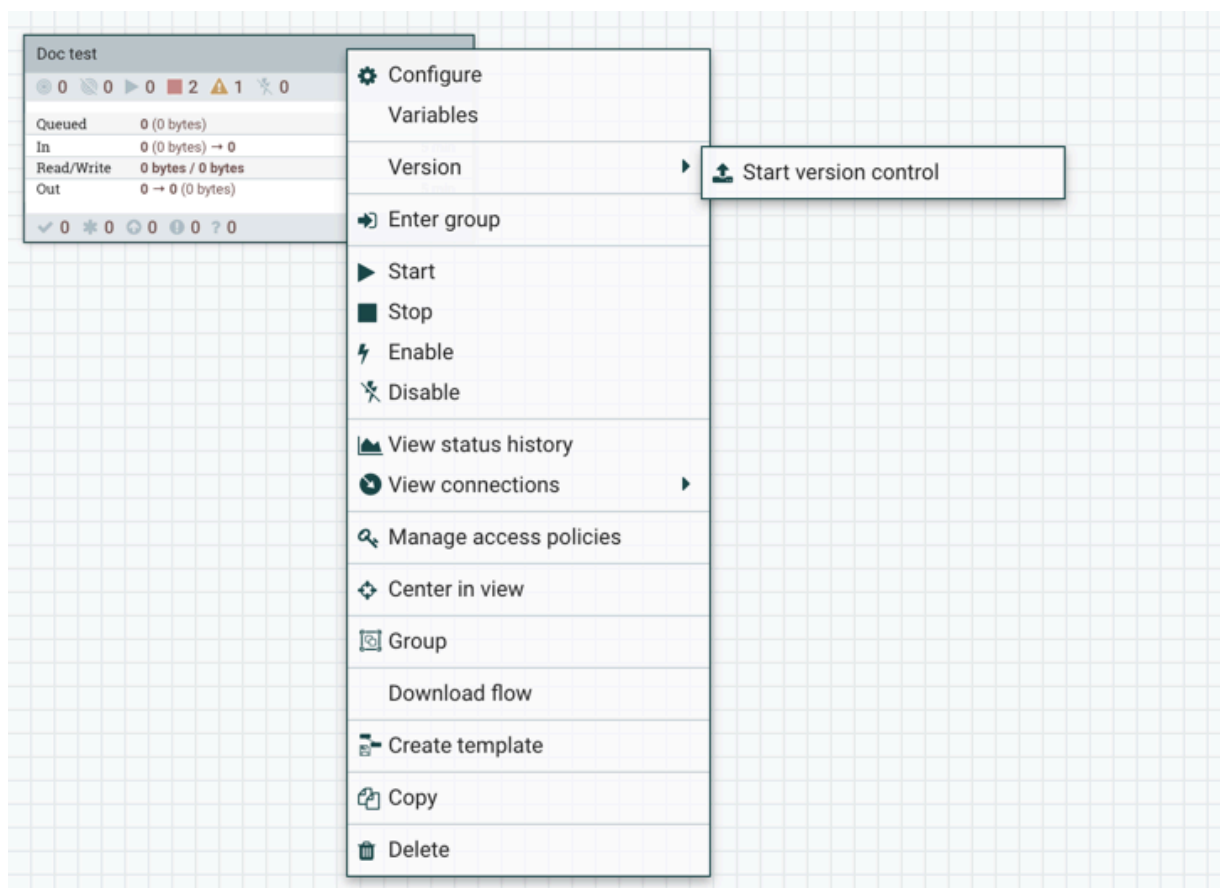
- You must have a process group already created in NiFi.
- You have connected NiFi to NiFi Registry.
- You have created a bucket in Registry for storing and organizing your NiFi data flows.

Start version control

Learn how you can place a process group under version control in NiFi.

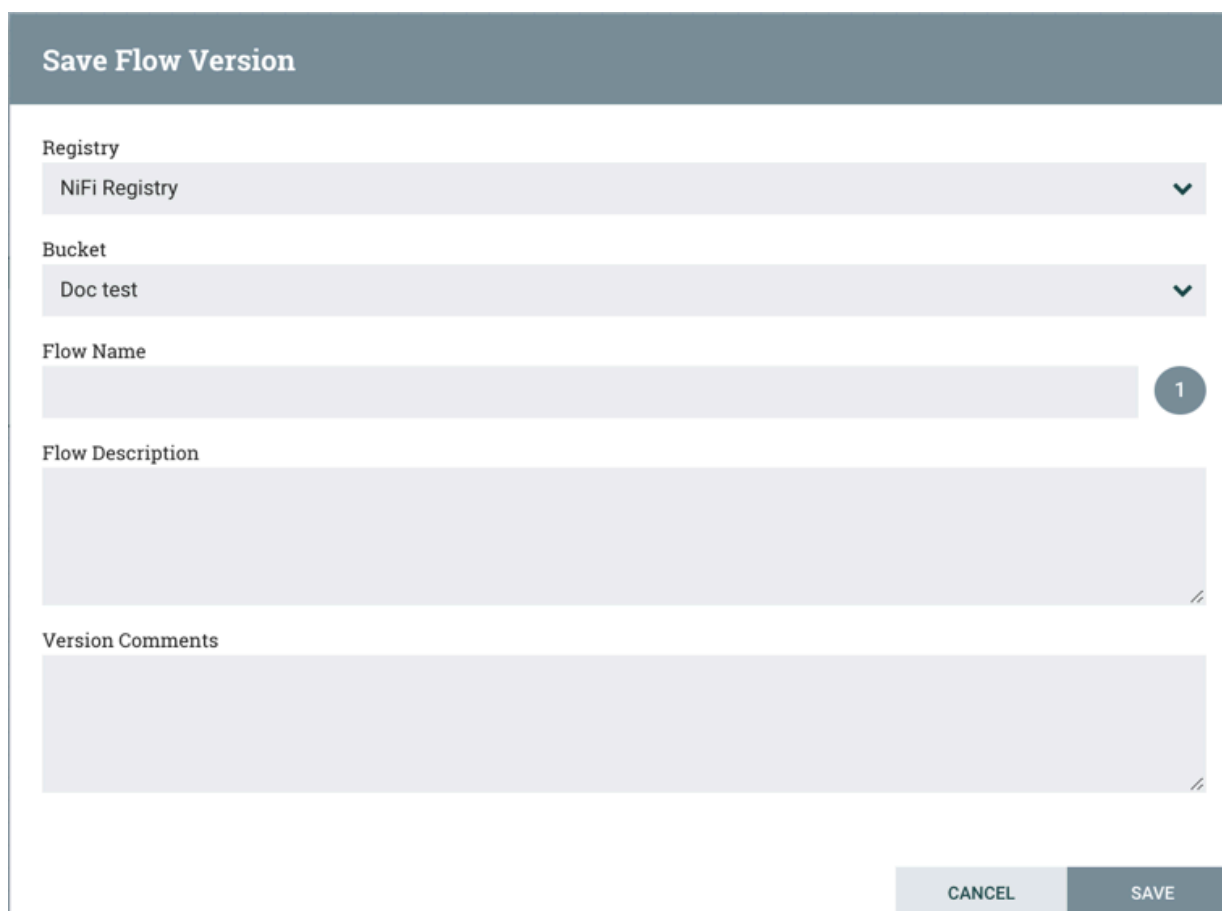
Procedure

1. Right-click the process group, then click VersionStart version control.



2. Specify the NiFi Registry bucket where you want to save your versioned data flow.

3. Set the flow name, flow description, and version comments.



The image shows a 'Save Flow Version' dialog box. It has a dark grey header with the title 'Save Flow Version'. Below the header, there are five sections: 'Registry' with a dropdown menu showing 'NiFi Registry'; 'Bucket' with a dropdown menu showing 'Doc test'; 'Flow Name' with a text input field and a circular icon containing the number '1'; 'Flow Description' with a large text area; and 'Version Comments' with a large text area. At the bottom right, there are two buttons: 'CANCEL' and 'SAVE'.

4. Click Save.

Results

Your flow version is saved and is ready to be deployed. The process group has a version state icon in the upper-left corner. If it is a green checkmark, it indicates that the process group is saved as a versioned flow in the Registry.

Commit local changes

Learn how you can manage local changes made to your versioned data flow in NiFi.

About this task

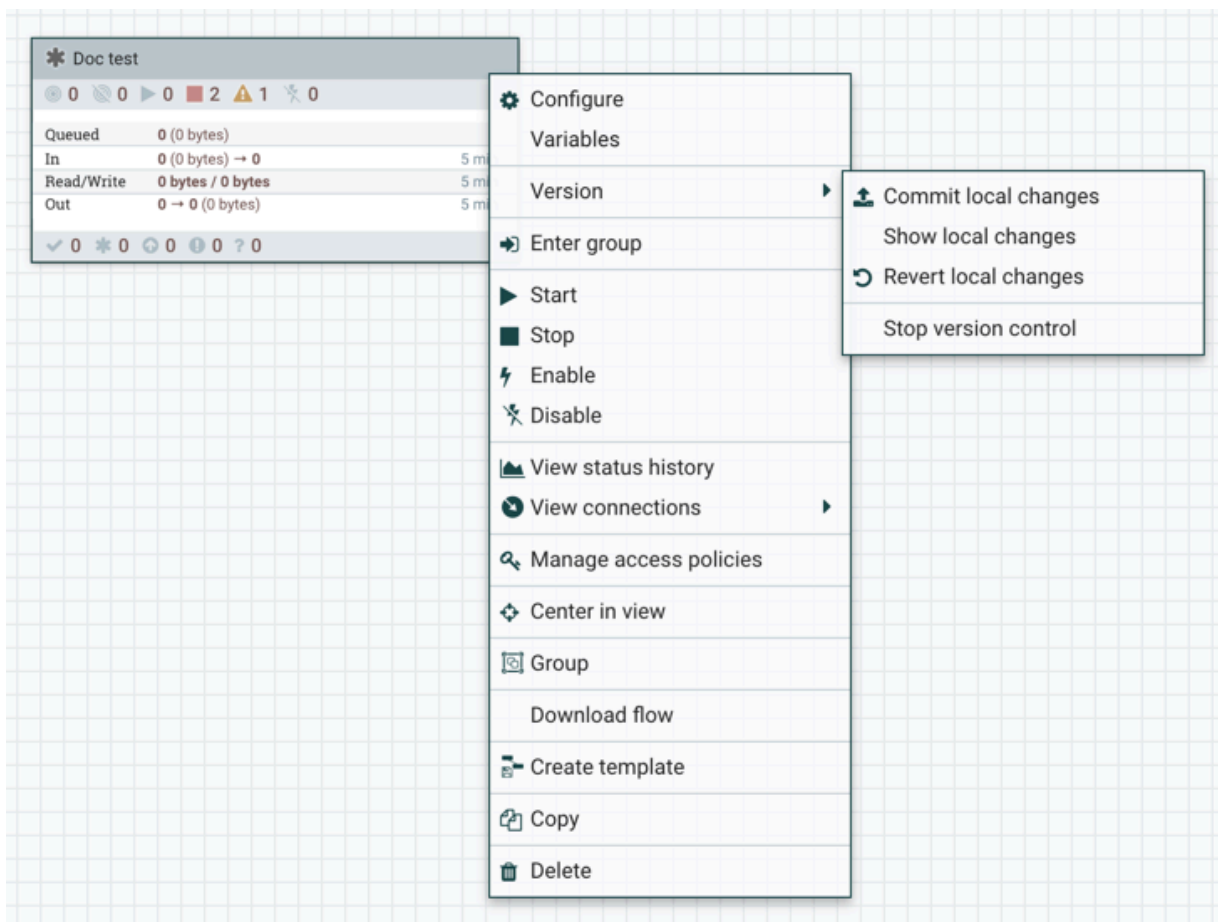
When you make changes in a data flow, the changes are local at first. If you want these changes to be saved as a separate flow version, you have to commit these changes to the NiFi Registry.

If you make changes to a versioned process group, the green checkmark icon changes to a grey asterisk because of the uncommitted local edits.

Procedure

1. Right-click the process group or the canvas in the workflow window and choose Version.

The context menu shows that you can save, review, or revert your changes.



- Select Commit local changes if you want to save the changes as a new version of the data flow.
 - Select Show local changes if you want to see the details of the edits.
 - Select Revert local changes if you want to discard the edits.
2. Select Commit local changes.
The Save Flow Version pop-up window appears, where you can add comments to the version to be saved.
 3. Click Save.

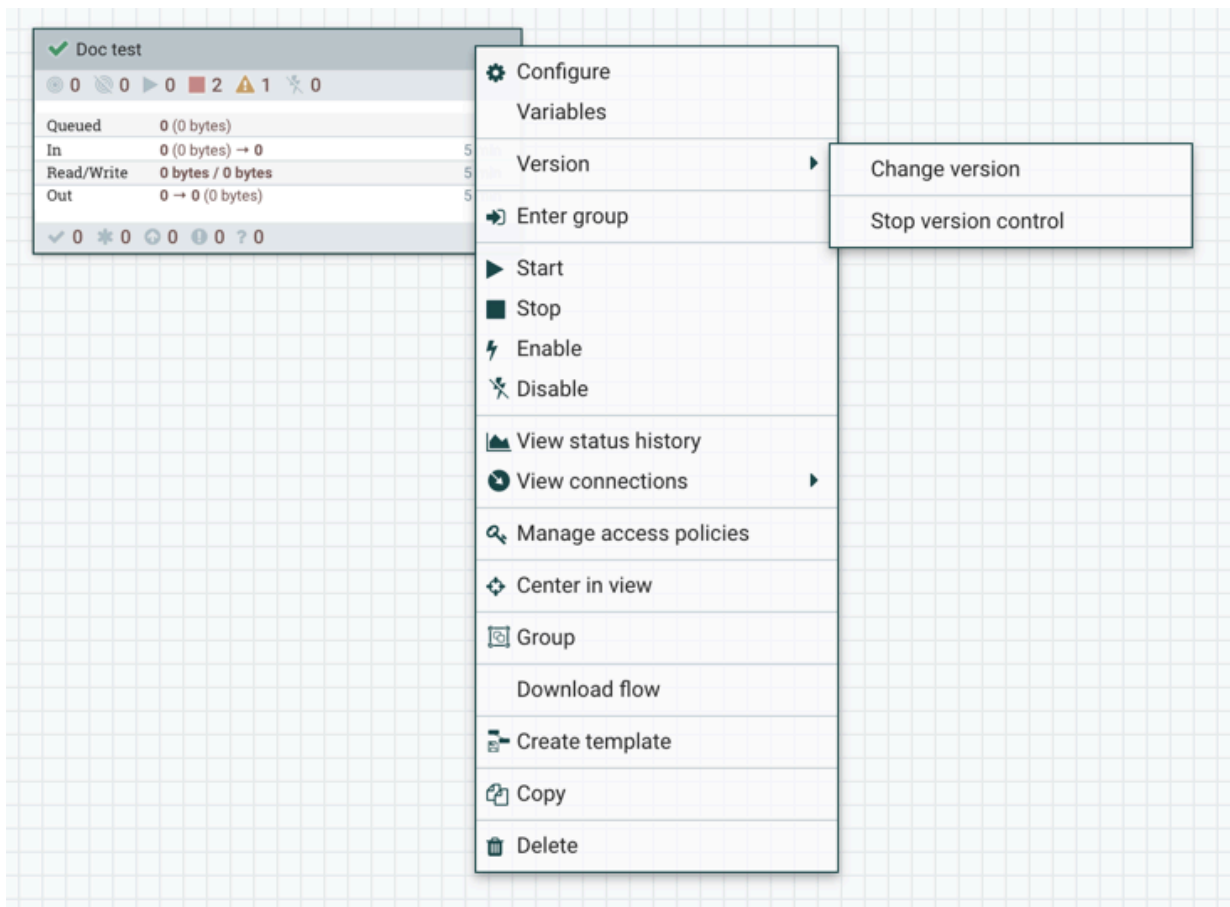
Change the version of your data flow

If you have multiple versions of a data flow saved in NiFi Registry, you can go back to any of the versions. Learn how you can change the versions of your data flows in NiFi.

Procedure

To open a different flow version:

1. In the NiFi UI, right-click the process group and go to VersionChange version to open the list of available versions.



2. Select the version.
3. Click Change.

Results

Your flow version is changed.

Manage versioned data flows

Learn how you can work with the data flows stored in your NiFi Registry buckets.

Procedure

- You can view your versioned flows in NiFi registry in a single list or by bucket.
- You can sort and filter the flows according to different criteria.
- You can also delete a versioned flow from the registry.
- You can check the changelog of each flow for details on its versions. You can get information on who made a change and when the change was made. If You can also take some actions on the version flow.

For detailed instructions on how to manage flows, see the *Apache NiFi Registry User Guide*.